Summary

- Motivation
- Goals
- The Metadata Advertising Protocol
- Proof of Concept
- Call for Feedback
- References

Motivation
What is this talk about?

- Data Retrieval problem.
    - Clients accessing remote data repositories.
- Directory Service Providers (DSPs)
    - White Pages, Libraries, Dictionaries, Data Repositories in general.
- Structured data in some way
    - Tagged information, records, fields, whatever.
- Data Retrieval vs. Information Retrieval

Motivation. Why?

- Client side (searching for data)
  - What DSPs may contain what I want?
  - How should I ask those DSPs?
- Directory Service Provide side (offering data)
  - How can I make the data I offer known to the public?
  - How can I make my data accessible in the same way as other DSPs offering the same kind of data?

Goals

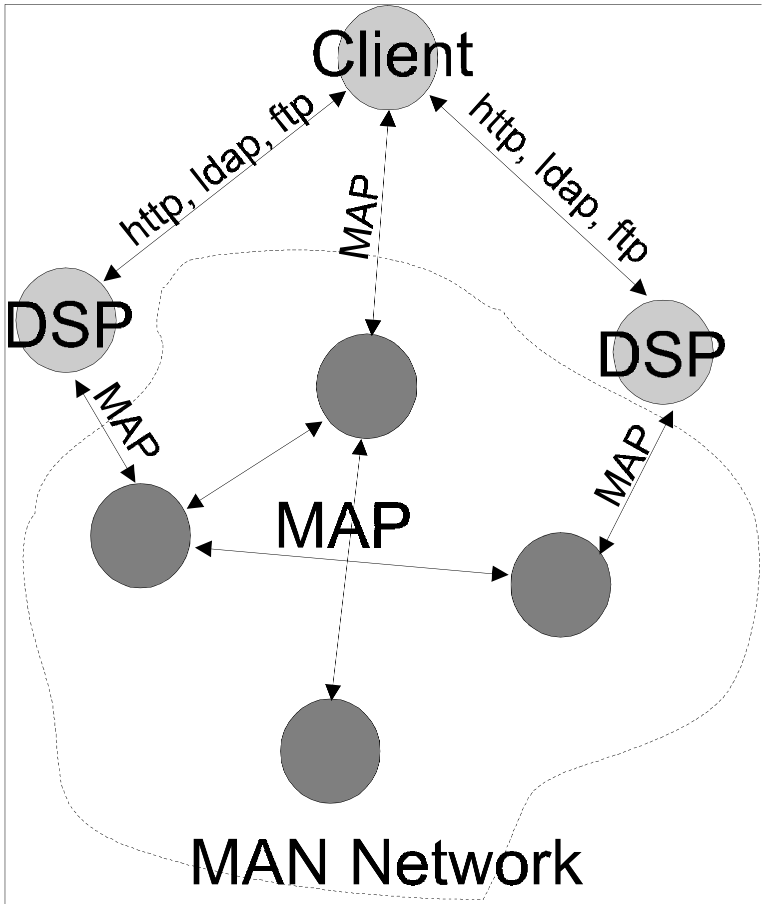
- To offer CLIENTS and DSPs a *common ground* where to *organise*, *advertise* and *navigate* through the information about the data globally available.
- To create a *forum* where DSPs advertise and organise their services using the same language.
- To *stimulate similar DSPs* at advertising their data in similar ways.

MAP: A Metadata Advertising Protocol
First Needs

- **A Metadata Language**
  - For DSPs to specify the data they contain.
  - To allow organisation and unification of DSPs specs.
- **A mechanism to make it Globally Available**
  - Anyone can access any piece of metadata.
- **A User Interaction Model**
  - To search, browse and navigate through the metadata.
  - To give the metadata a practical value.

MAP: A Metadata Advertising Protocol
Governing Mechanisms

- **Global metadata is maintained with an NNTP-like mechanism**
- **DSPs advertise in any MAN**
- **Clients navigate metadata in any MAN**
- **MANs search the metadata for possible similarities among DSPs**



MAP: A Metadata Advertising Protocol
Querying DSPs

- DSPs introduce an specification of their services in the metadata.
- Clients use the specification of a DSP to query it in its own terms.
- Thus, DSPs don't have to adapt their already existing data to any standard.
- Client can make distributed queries by using specifications of several DSPs

Proof of Concept
A Prototype Implementation

- **Metadata. Three kinds of entities**
  - **Categories, Object Types and DSPs**
- **Prototype Server. Three simple commands**
  - **To match against entity types.**
  - **To search.**
  - **To get new entities.**
- **User Interaction Model. Two ways**
  - **Navigation through an organised tree.**
  - **Search the entire metadata.**

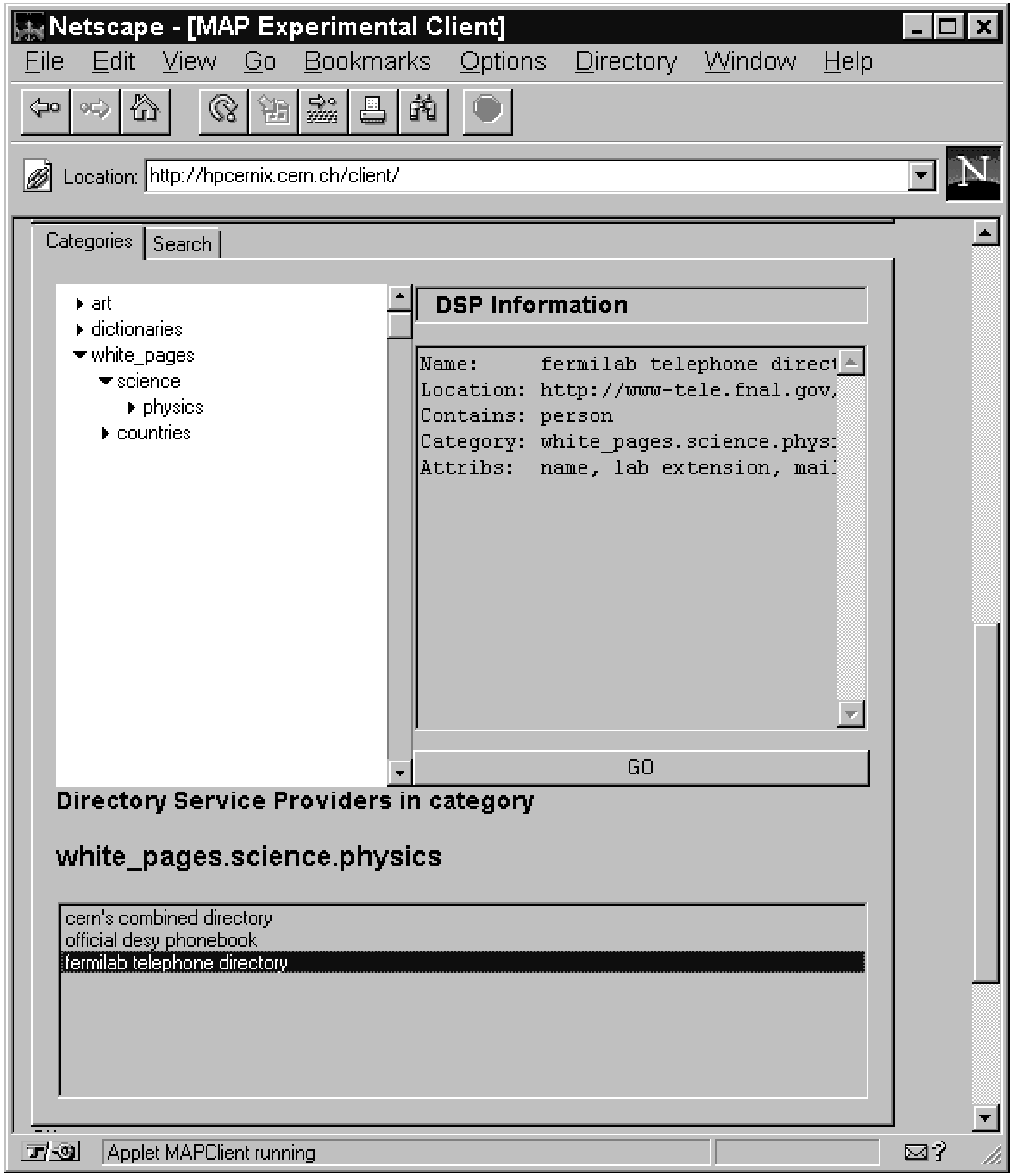
Proof of Concept
An Example of Metadata

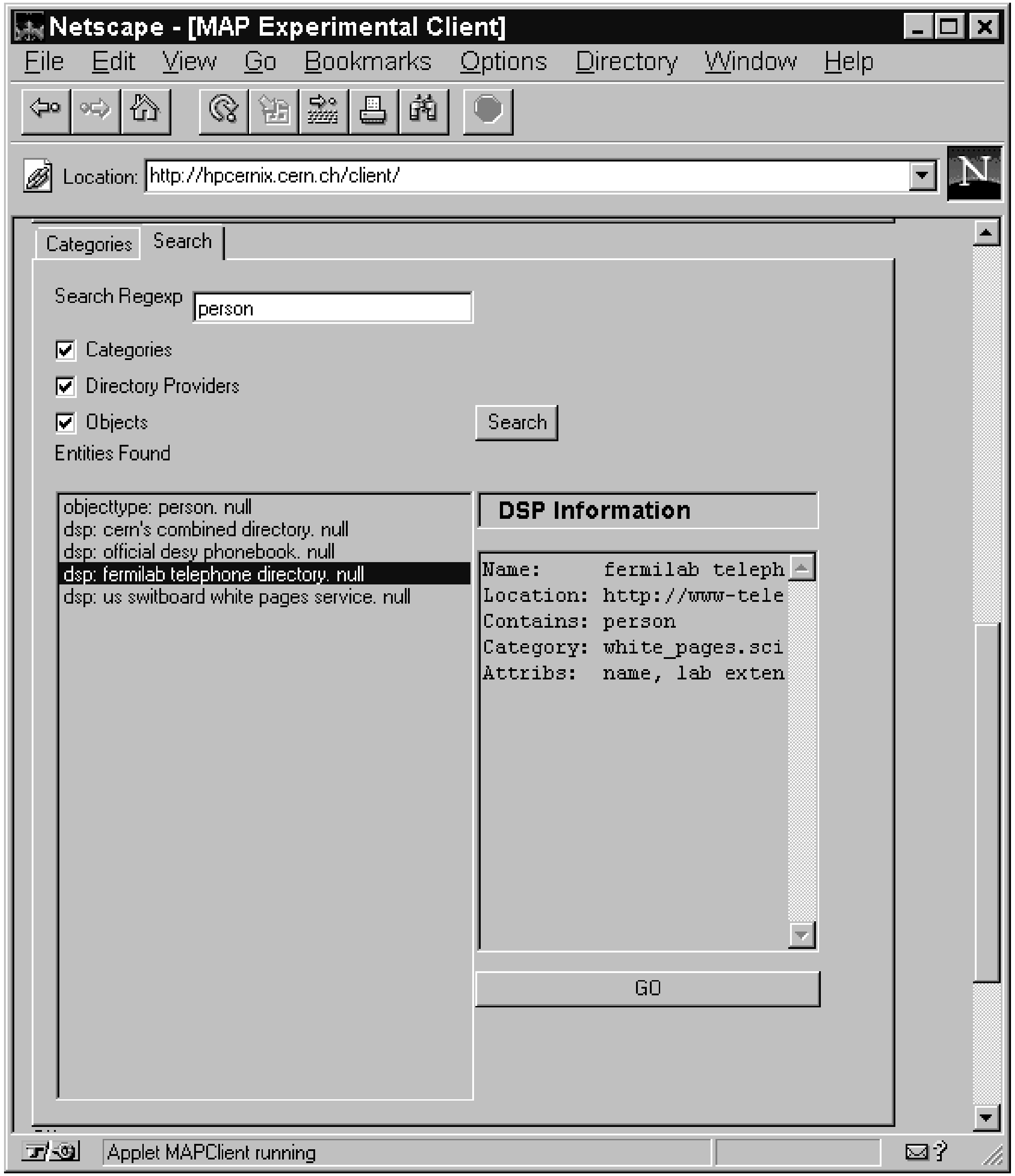
```
Entity:         dsp
Name:           CERN's combined directory
Location:       http://consult.cern.ch/xwho
Admin:          Phone.Master@cern.ch
Objecttype:     person
Status:         not classified, object type normalised
Category:       white_pages.science.physics
Attributes:     firstname, surname, email, phone, beep, division, building

Entity:         category
Name:           white_pages
Parent:         .

Entity:         category
Name:           countries
Parent:         white_pages

Entity:         objecttype
Name:           person
Attributes:     firstname, family name, address, phone, email
Synonyms:       surname:family name, telephone:phone
Content:        people, individuals, particular
```


Netscape - [MAP Experimental Client]
File Edit View Go Bookmarks Options Directory Window Help
Location: http://hpcernix.cern.ch/client/
Categories
Search
art
dictionaries
white_pages
science
physics
countries
DSP Information
Name: fermilab telephone direct
Location: http://www-tele.fnal.gov,
Contains: person
Category: white_pages.science.phys
Attribs: name, lab extension, mai
GO
Directory Service Providers in category
white_pages.science.physics
cern's combined directory
official desy phonebook
fermilab telephone directory
Applet MAPClient running


Netscape - [MAP Experimental Client]
File Edit View Go Bookmarks Options Directory Window Help
Location: http://hpcernix.cern.ch/client/
Categories
Search
Search Regexp
person
Categories
Directory Providers
Objects
Search
Entities Found
objecttype: person. null
dsp: cern's combined directory. null
dsp: official desy phonebook. null
dsp: fermilab telephone directory. null
dsp: us switboard white pages service. null
DSP Information
Name: fermilab teleph
Location: http://www-tele
Contains: person
Category: white_pages.sci
Attribs: name, lab exten
GO
Applet MAPClient running

Current Work

- Stablish a Metadata language
- Stablish a Server specification

Having in Mind

- Consider reusing existing technology
- A practical approach may be a simplified version of a theoretical one.
- Once running new DSPs are easily added

For more information...

- **MAP Project WEB page**

  `http://wwwmap.cern.ch`

  - **Project Documentation.**
  - **Prototype Server and Client application (java).**

- **MAP Mailing List**

  `map-forum@listbox.cern.ch`

Theoretical Background

- **The Metadata Language**
  - **Description Logics and Terminological Knowledge Representation Systems**
  - **A full implementation would allow distributed reasoning based on data stored on different DSPs**
- **Metadata Analysis Mechanisms**
  - **Clustering, Machine Learning, Knowledge Discovery Techniques**
  - **Different nodes can have different analysis approaches**

A Taxonomically Distributed Data Retrieval Model

M.Marquina, R.Ramos, A.Taddei
CERN / IT Division